MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

Bolt Beranek and Newman Inc.

bbn

A083080 (1)

Report No. 4384
Contract No. MDA940-79-C-0411

LEVEL #5

AD**A085972**

EFFICIENT ENCODING AND DECODING OF SPEECH

Quarterly Progress Report No. 2
1 February 1980 - 30 April 1980

JUN 2 3 1980

A

Submitted to:

Mr. David Fonseca, R814
9800 Savage Road
Fort George G. Meade, MD 20755

80 6 19 074

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER BBN Report No. 4384	2. GOVT ACCESSION NO. AD-A085972	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) EFFICIENT ENCODING AND DECODING OF SPEECH		5. TYPE OF REPORT & PERIOD COVERED 2nd Quart. Progress Report Feb. 1980 - April. 1980
		6. PERFORMING ORG. REPORT NUMBER Report No. 4384
7. AUTHOR(s) M. Berouti J. Makhoul M. Krasner		8. CONTRACT OR GRANT NUMBER(s) MDA904-79-C-0411
9. PERFORMING ORGANIZATION NAME AND ADDRESS Bolt Beranek and Newman Inc. 50 Moulton St. Cambridge, MA 02138		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Maryland Procurement Office 9800 Savage Rd., Ft. George A. Meade, MD 20755		12. REPORT DATE May 1980
		13. NUMBER OF PAGES 36
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce for sale to the general public.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Speech compression, linear prediction, adaptive predictive coding, down-samppling, quantization, pitch filter, preemphasis, noise shaping, variable-rate output, filter design.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We report on continued research to improve the APC algorithm. The issue of system stability as a function of the linear prediction filter is discussed and a method for improving performance presented. The issue of computational complexity is also addressed. A new filter that reduces the computational load of the algorithm is designed for use in the resampling operations.

DD FORM 1473 1 JAN 73   EDITION OF 1 NOV 65 IS OBSOLETE

Report No. 4384

EFFICIENT CODING AND DECODING OF SPEECH

Quarterly Progress Report No. 2
1 February – 30 April 1980

Prepared by:

Bolt Beranek and Newman Inc.
50 Moulton Street
Cambridge, Massachusetts   02138

Prepared for:

Mr. David Fonseca, R814
9800 Savage Road
Fort George G. Meade, MD 20755

TABLE OF CONTENTS

1. INTRODUCTION

This report describes work performed during the second quarter of our contract for the efficient encoding and decoding of speech. Our work during this quarter concentrated on two areas basic to the project: the reduction of the amount of computation in processing speech, and the determination and removal of the causes of very large errors for some speech frames.

It was noted in the last progress report that large coding errors occur for some frames of speech. This problem can be related to the basic APC system algorithm as a function of the linear prediction filter. In Section 2 of this report, the problem is studied for cases that represent typical APC systems and a method of optimization of the coding is presented.

As also noted in our last report, the vast majority of computations involved in the APC algorithm occurs during the resampling operations. This is due to the long FIR lowpass filters that are used to reduce the aliasing that occurs during resampling. During this quarter, we studied the problem of filter design for resampling based on the specific properties of the speech signal. A new filter that requires much less computation has been designed and is presented in Section 3.

The report concludes in Section 4 with a description of our plans for research during the next two quarters of the project.

2. ANALYSIS OF THE APC FEEDBACK LOOP

An often used and quite accurate approximation to the signal to quantization noise ratio, S/Q, of an APC system is the product of the linear prediction gain, S/R or V_p^{-1}, and the quantizer input to quantization noise energy ratio, W/Q. When noise spectral shaping is employed, the noise energy is increased, lowering the overall S/Q, while the geometric mean of the noise spectrum stays constant. Often, however, the APC output contains frames with S/Q that are much smaller than expected. For APC systems both with and without noise shaping and using quantization that is matched to either fixed-length or variable-length codes, some frames are perceived as "glitches" or "beeps" and have corresponding S/Q that are less than unity, i.e., negative in dB. This is indicative of the noise energy being larger than the speech signal energy.

Although the autocorrelation method of linear prediction guarantees that the all-pole filter is stable, the stability of the APC feedback loop can not be guaranteed in general. As the loop contains a non-linear element, the quantizer, it is not possible to analyze the response of the system for arbitrary inputs. It is possible to gain some insight by making some reasonable assumptions and applying classical techniques.

This section will analyze the APC system for quantizers with a small number of levels that are matched to fixed-length coding schemes and for quantizers with a large number of levels that are matched to variable-length, entropy coding schemes. Although the analysis is given in terms of the APC system without spectral noise shaping, the results are easily applied to the noise shaping case. The result of the analysis is an understanding of the cause of instability in the APC loop. The method we reported preciously to eliminate the degradations due to these instabilities is improved.

2.1 Equivalence of System Configurations

Two configurations that are used to implement the APC system are shown in Figs. 1 and 2. It is easily shown that the outputs of each of these systems are the same. It is not obvious, however, that the feedback loops for each of these systems are identical and will have identical properties even when unstable. This is made clear by redrawing the loop as in Fig. 3. The only difference between the two configurations is the place at which the input occurs into the loop. In Fig. 3, the upper and lower positions of the input switch correspond to Figs. 1 and 2, respectively.
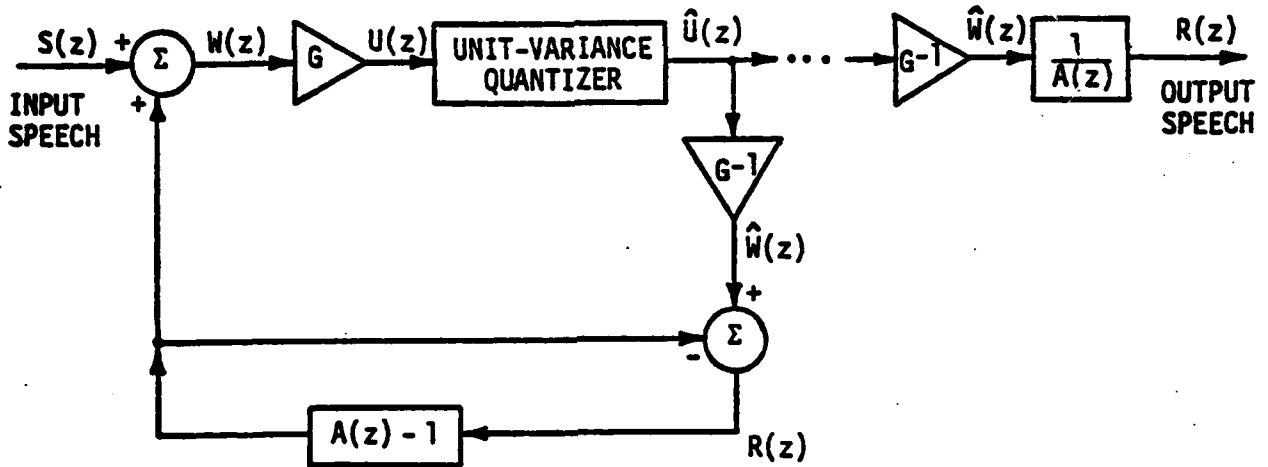
FIG. 1. PREDICTION FEEDBACK CONFIGURATION
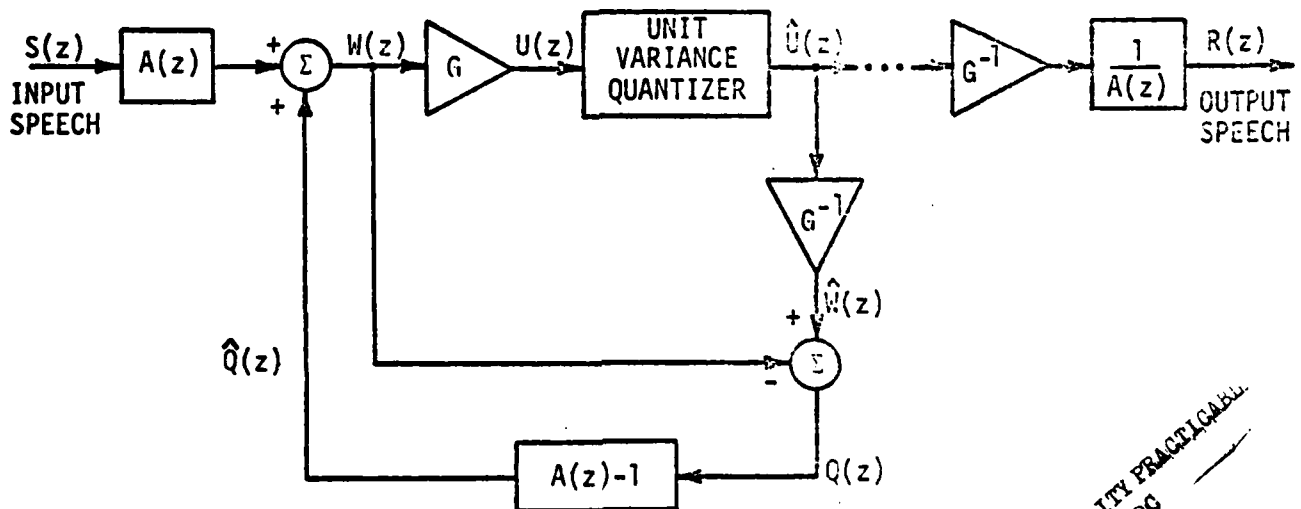
FIG. 2. NOISE FEEDBACK CONFIGURATION
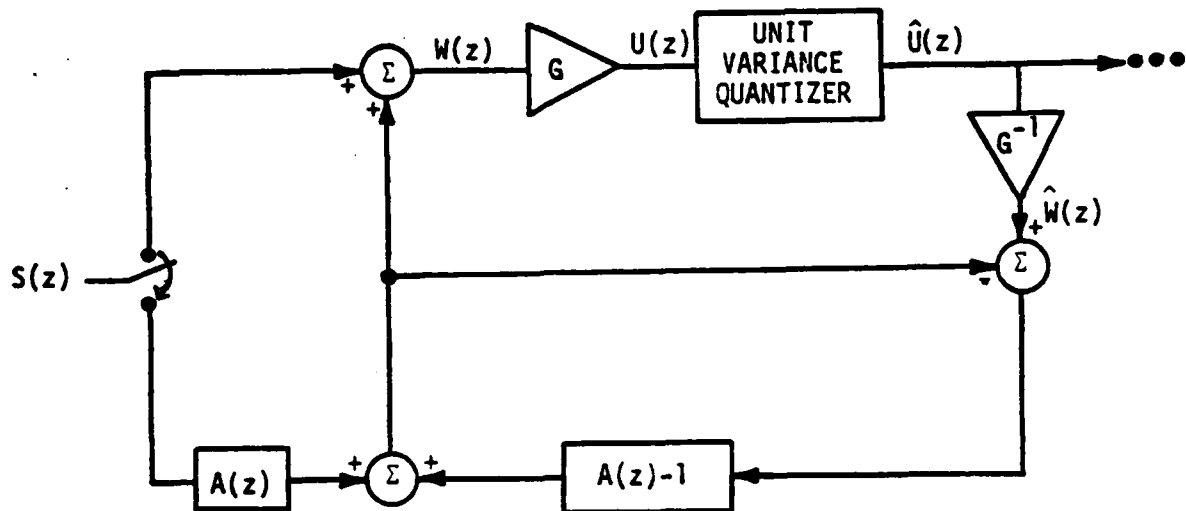
FIG. 3. FEEDBACK LOOP OF FIGS. 1 & 2

Since the configurations produce the same output, either one can be chosen for the analysis. The parametric analysis of the stability properties of the system is facilitated by looking at the noise feedback configuration of Fig. 2.

The analysis is dependent upon the properties of the quantizer within the loop. As the nonlinear input-output relation for a quantizer that is matched to a variable-length coding scheme is very different from that for a quantizer that is matched to a fixed-length coding scheme, the analysis for each case will be treated in a separate section.

2.2 Quantization for Entropy Coding Schemes

The quantizer that is matched to a variable-length, entropy code has a large number of levels. For the purpose of analysis, the number of levels is assumed to be infinite. In practice, this is not necessary as the probability of occurrence of all but a small number of the levels is negligible. We have found that for quantization using approximately 2 bits per sample, 23 levels is sufficient.

In Fig. 4, the noise feedback configuration is presented in a parametric representation using the variance (or power) of the signals in the APC loop as the parameters. Specifically, σ_s^2 is the input speech variance, σ_e^2 is the linear prediction residual variance, σ_w^2 is the quantizer input variance, and σ_q^2 is the quantization error variance. For the purposes of the feedback loop analysis, the quantizer can be represented by a linear gain block with the gain being the ratio of the quantization error power to the quantizer input power. The inverse of this quantity is defined in this report as the signal to noise ratio of the quantizer, W/Q. This gain, $(W/Q)^{-1}$ is dependent on the power in the quantizer input and on the spacing of the quantizer threshold levels. If that spacing is determined as a function of the quantizer input power, the gain of the quantizer block can be

forced to a constant. For any fixed distribution of quantizer input amplitudes, this will also force the entropy of the quantized signal to a constant. This idea is basic to the variable-to-fixed rate conversion scheme reported in the last report.

A different algorithm is to set the spacing as a function of the LPC residual power such that the residual to quantization noise power ratio, R/Q, is constant. The entropy of the quantized signal and, therefore, the bit rate of the system, will not be constant. The implications of these two schemes are discussed in the following sections.

For the purpose of the analysis, the following assumptions are made: the quantization error can be modeled as a white noise process; the quantization error is statistically independent from the input speech signal; and the amplitude distribution of samples at the input to the quantizer does not change as as function of the gain of the feedback filter.

2.2.1 Quantizer Matched to Residual Energy

The first case to be analyzed is for a quantizer that is matched to the energy in the LPC residual, i.e., the signal at the input to the feedback loop in Fig. 2. This very simple
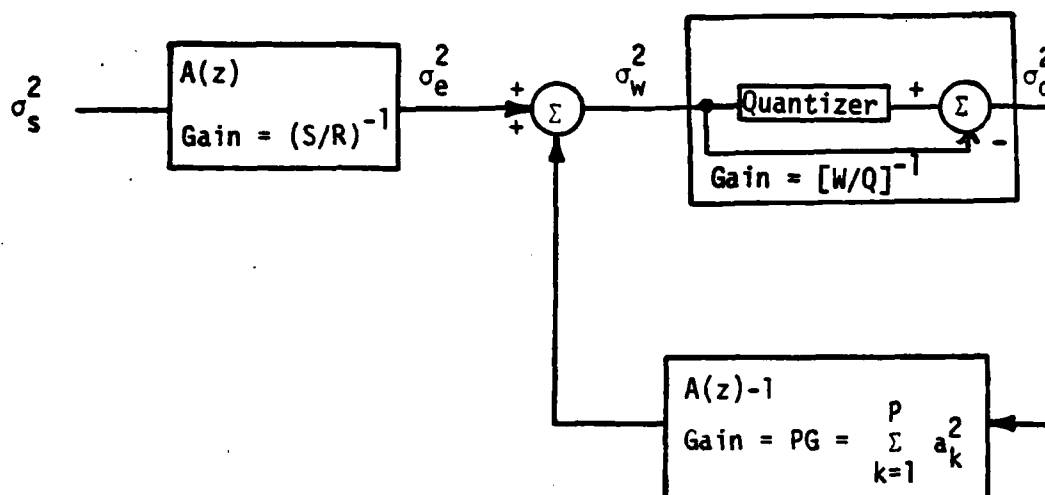
FIG. 4. PARAMETRIC REPRESENTATION OF THE APC LOOP

scheme was used in the APC system before the iterative variable
to fixed rate algorithm was formulated. For this case, the
residual to quantization noise ratio, R/Q, is forced to be
constant.

The matching or normalizing of the quantizer is the setting
of the quantizer threshold levels such that for the quantization
of that signal, the entropy, and, hence, the bit rate needed for
coding, are fixed to required values. If another signal is the
actual input to the quantizer, the entropy of the quantized

9

signal will change.   The quantizer error power, primarily determined by the quantizer level spacing, will change little.

When the quantizer is matched to a specific input signal such that the average code length is 2.12 bits per sample (as in this APC system), the ratio of the powers of that signal to the quantization error is approximately 11 dB, varying as a function of the exact signal amplitude distribution.   Thus, for a quantizer matched in this manner to the residual power, the R/Q is approximately 11 dB.   The S/Q of a frame will be the S/R power ratio of that frame, approximated by $1/V_p$, multiplied by R/Q.

The variance of the quantization noise is primarily a function of the spacing of the quantization threshold levels. For a given amplitude distribution of the quantizer input signal, the spacing will determine both the entropy of the quantized samples and W/Q. If the quantizer is matched to the power of the LPC residual such that R/Q is fixed independent of the quantizer input power, the entropy of the quantized signal and, hence, the bit rate needed to code the signal will be a function of the residual-to-quantizer-input ratio, R/W.   This can be shown to be a function of the power gain, PG, of the feedback filter, A(z)-1.

Under the assumptions mentioned earlier, the power of the

10

input to the quantizer is given by

$$\sigma_w^2 = \sigma_e^2 + \sigma_q^2 \sum_{k=1}^{P} a_k^2$$

$$= \sigma_e^2 + \sigma_q^2 \, PG$$

(1)

where σ_e^2 is the variance of the LPC residual, σ_w^2 is the variance of the input to the quantizer, σ_q^2 is the variance of the quantization error, and PG is the power gain of the feedback filter. For a quantization noise with a flat spectrum, that power gain is just the sum of the squares of the coefficients, i.e., the energy in the impulse response. Since the quantizer input has a power larger than the signal that the quantizer is matched to, i.e., R/W is less than unity, the output entropy will be increased over the value used in setting the quantization levels. For arbitrary distributions, the resultant minimum coding rate can not be calculated in a closed form. If the distribution is Gaussian, then rate distortion theory predicts that the increase in the minimum required coding rate is given by

$$R_w(d) = R_e(d) + \frac{1}{2} \log \frac{\sigma_w^2}{\sigma_e^2} \tag{2}$$

where $R_w(d)$ and $R_e(d)$ are the minimum rates per sample necessary for the coding and quantization with a mean square distortion of d given by rate distortion theory for signals with Gaussian distribution. This rate assumes the use of an optimal coding scheme which is not specified. Although the rate distortion function varies for each distribution, this equation may be used as an approximation to the minimum bit rate required for coding of signals with non-Gaussian distributions.

Assuming an R/Q of 11 dB in (1), a PG of 16 dB would yield a quantizer input to residual power ratio of 4. An optimum coding scheme then would require an additional 1 bit per sample for coding.

We conclude this subsection by pointing out that in the variable rate APC system where the quantizer is matched to the LPC residual power, the bit rate varies from frame to frame. These variations in bit rate are primarily governed by the effect of PG on the power to the input to the quantizer, as explained above.

2.2.2 Quantizer Matched for Constant Bit Rate

The second case of interest is a system in which the output bit rate is a constant. For many applications, a specified constant coding rate is necessary because of communication channel requirements. Although a constant coding rate for each sample, i.e., a fixed length code, is not often necessary, the number of bits used to encode every interval of some fixed time interval may be required. The system using the speech coder would then buffer the coded samples for that time interval. A typical time interval for that buffering would be 20 to 40 ms corresponding to 1 to 2 frames of speech in the APC system. Thus, variable length coding schemes can be used for many more applications if the coding rate can be forced to a constant value over given small intervals of time. This section discusses the process of matching the quantizer to the quantizer input. This will force the average bit rate to be constant. The problem of making the number of bits for encoding a frame exactly equal to a fixed number is not discussed.

For the average bit rate to be a constant value, the quantizer threshold levels must be a function of the quantizer input power and distribution. Then, the S/Q of the system will be less than the case of the last section where the quantizer was

matched to the LPC residual. This decrease in S/Q is equal to the factor R/W. When the bit rate is forced to be a constant, the W/Q is approximately constant and the loop must be analyzed as a feedback system. From (1), the quantization error to residual power transfer function, R/Q^{-1}, is given by

$$[R/Q]^{-1} = \frac{\sigma_q^2}{\sigma_e^2} = \frac{\sigma_q^2 / \sigma_w^2}{1 - \frac{\sigma_q^2}{\sigma_w^2} \sum_{k=1}^{P} a_k^2}$$

$$= \frac{[W/Q]^{-1}}{1 - \frac{PG}{W/Q}}$$

$$= \frac{1}{W/Q - PG} \tag{3}$$

and the signal to output noise ratio, S/Q, is then

$$S/Q = S/R \cdot R/W \cdot W/Q$$

$$= \frac{\sigma_s^2}{\sigma_e^2} (1 - \frac{PG}{W/Q}) \; W/Q$$

$$\approx V_p^{-1} (W/Q - PG) \tag{4}$$

where a_k are the coefficients of the filter A(z), PG is the power gain of the feedback filter A(z)-1, W/Q is the quantizer input to quantization noise ratio, V_p^{-1} is the prediction gain found by the autocorrelation method of linear prediction used in the system, and R/W is given by

$$R/W = \frac{\sigma_e^2}{\sigma_w^2} = 1 - \frac{PG}{W/Q} \qquad (5)$$

The system signal-to-noise ratio, S/Q, is a function of the prediction gain and the power gain. By matching the quantizer for a constant bit rate, W/Q is constant. Assuming that V_p is constant, S/Q is proportional to R/W. To show how S/Q is affected by PG, R/W, as in (5), is plotted in Fig. 5 as a function of PG. In Fig. 5, PG is shown in dB relative to the value of W/Q which is assumed independent of the filter. Note that when PG approaches W/Q, R/W decreases infinitely indicating an infinitely growing quantizer input.

When the power gain, PG, is much smaller than W/Q, the feedback term into the summation block of the feedback loop is negligible. The S/Q is then the product of the prediction gain
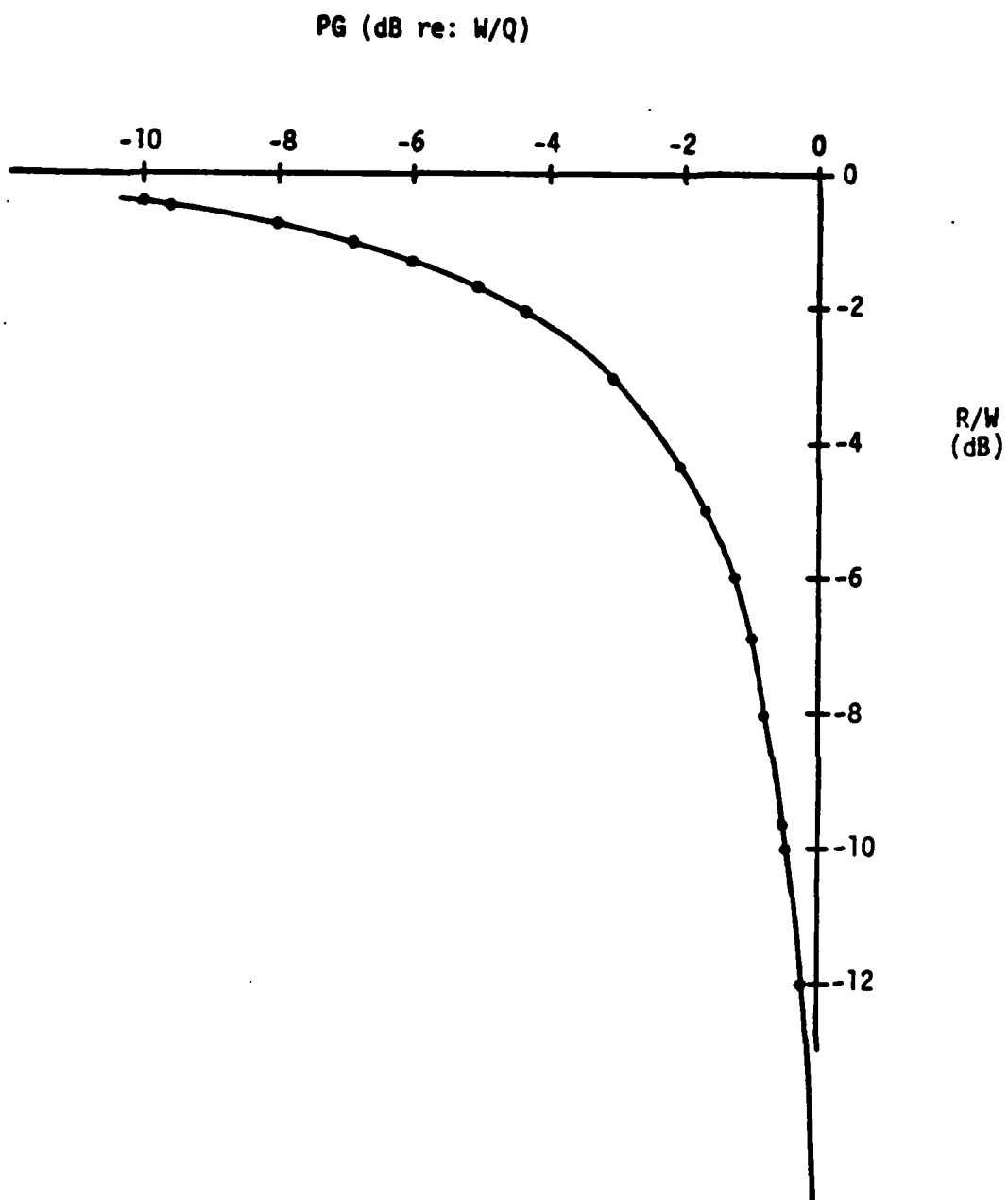
PG (dB re: W/Q)

FIG. 5. RESIDUAL TO QUANTIZATION INPUT POWER RATIO, R/W

and W/Q. When the power gain is of the same order of magnitude as W/Q, it will reduce the S/Q since the feedback term to the summation block is of the same order of magnitude as the residual. When PG equals W/Q, R/W^{-1} is infinite. Since W/Q is constant, the quantization error is infinite. When PG is greater than W/Q, the equation predicts that the ratio R/W is negative. This last case corresponds to an unstable system as the ratio of two positive quantities, the powers, can never be negative. A correct interpretation is that the assumption that W/Q is constant is not true: there is no spacing of quantizer threshold levels that will yield the required bit rate and, therefore, a value of W/Q. Any chosen spacing will cause the power in the feedback term to be too large for the required bit rate.

In terms of the APC system, this explains the failure of the variable to fixed rate conversion scheme to converge for some frames as described in the last progress report. For some feedback filters, the power gain is so large that no quantizer will yield the required bit rate. (An exception is, of course, the case where all samples are quantized to zero. This will have an entropy of zero or, using the fixed self-synchronized code, the minimum code length of one bit per sample).

17

2.3 Quantization for Fixed-Length Coding Schemes

The analysis for a quantizer with a small number of levels and followed by a fixed-length coding scheme, differs from the above in that the number of bits required for coding is independent of the signals in the APC loop. The bit rate depends only on the number of levels used in the quantization.

Although an optimal coding scheme would require a different number of bits due to the change in the entropy of the quantized signal, the fixed-length coding scheme will use the same number of bits independent of the distribution. Given the fixed-length coding scheme, an optimal choice of the quantizer threshold level spacing will minimize the mean square error and, hence, maximize the S/Q.

The S/Q of the system can still be calculated from (4). The performance of the quantizer as measured by W/Q is a function of the spacing of its threshold levels. Since any chosen spacing will determine the quantization noise and, therefore, will affect the quantizer input level, the optimization of the quantizer must account for the performance of the feedback loop. An iterative procedure can be used for this purpose.

For the fixed-length coding scheme, W/Q will attain its

maximum value when it is matched to the actual quantizer input signal. From (4) and assuming that the filter $A(z)$ has been calculated such that V_p and PG are fixed, S/Q is a monotonically increasing function of W/Q. Thus, the maximum S/Q of the system can be realized only for a maximum W/Q, i.e., when the quantizer is matched to the actual quantizer input power. Operation with other quantizer threshold values will be suboptimal. A design procedure would be to estimate the W/Q given the number of bits per sample for quantization assuming that the quantizer is matched to its input. The quantizer input power can then be calculated from the residual power and the W/Q. This allows the quantizer threshold levels to be calculated such that the W/Q is consistent with its assumed value. The S/Q will then be as in (4).

2.4 Interpretation of Power Gain

In the previous sections, it has been shown that the performance of the APC system is dependent on the value of the power gain, PG, of the filter in the feedback loop, $A(z)-1$. For an APC system with either fixed-length coding or entropy coding adjusted for constant output bit rate, the S/Q is a monotonically decreasing function of PG. For an entropy coding system with quantization threshold levels adjusted to the residual power, the

coding bit rate is a monotonically increasing function of PG. It is not true, however, that smaller values of PG will yield improved system performance. Both PG and S/R are functions of the filter A(z). In general, small values of S/R occur concurrently with small values of PG. Thus, for the frames of speech where the optimal filter yields a large value of PG, the reduction of that PG by modification of the filter may or may not improve system performance. The interpretation of the power gain given in this section is useful for determining when and how to reduce the PG.

The normalized prediction error, V_p, is a monotonically decreasing function of p, the order of the linear prediction filter. As the order becomes infinite, V_p asymptotically approaches a minimum value, denoted as V_{min}, and the spectrum of $A^{-1}(z)$ becomes equal to the input speech spectrum S(z). This section assumes an infinite order predictor for the calculations of prediction gain and power gain.

V_{min}, the inverse of the maximum normalized linear prediction gain, can be calculated as the ratio of the geometric mean to arithmetic mean of the spectrum of the input signal as in (7).

$$P_k \overset{\Delta}{=} \left| S\left(\frac{j2\pi k}{N}\right) \right|^2$$

$$V_{min} \overset{\Delta}{=} \frac{E_{min}}{R_0} \tag{6}$$

$$= \frac{\left(\prod_{k=0}^{N-1} P_k\right)^{1/N}}{\frac{1}{N} \sum_{k=0}^{N-1} P_k} \tag{7}$$

$$\log V_{min} = \log E_{min} - \log R_0$$

$$= \frac{1}{N} \sum_{k=0}^{N-1} \log P_k - \log\left(\frac{1}{N} \sum_{k=0}^{N-1} P_k\right) \tag{8}$$

where P_k is defined as the magnitude squared of the input speech signal spectrum $S(\omega)$, evaluated at the frequency $\omega = (j2\pi k/N)$. V_{min} is the minimum value of the normalized error or, equivalently, V_{min}^{-1} is the maximum value of the prediction gain. By Parseval's theorem, the power gain and prediction gain are related to the whitening filter, $A(z)$, by assuming a finite time

window and a DFT representation of the spectra.

$$\frac{E(z)}{S(z)} = A(z) = 1 + \sum_{k=1}^{P} a_k \, z^{-k} \tag{9}$$

$$\frac{1}{N} \sum_{k=0}^{N-1} \frac{|E_k|^2}{|S_k|^2} = 1 + \sum_{k=0}^{N-1} a_k^2$$
$$= 1 + PG \tag{10}$$

So far, no information about the filter $A(z)$ has been necessary. This filter is the result of the linear prediction analysis and has a spectrum that is an approximation to the inverse of the speech spectrum. Thus, $A(z)$ is a whitening filter and the spectrum of $E(z)$ is nearly flat. Assuming

$$|E_k|^2 = E_{min} \, , \, \forall \, k \tag{11}$$

then from (10),

$$PG + 1 = E_{min} \cdot \frac{1}{N} \sum_{k=0}^{N-1} P_k^{-1} \tag{12}$$

$$\log (PG + 1) = \log E_{min} + \log \frac{1}{N} \sum_{k=0}^{N-1} P_k^{-1} \tag{13}$$

Let us define R_0' as the energy in the signal that has the inverse spectrum of the speech. Both the prediction gain and the power gain can then be written in a simple form.

$$R_0' \triangleq \frac{1}{N} \sum_{k=0}^{N-1} P_k^{-1} \tag{14}$$

$$\log V_{min} = \log E_{min} - \log R_0 \tag{15}$$

$$\log (PG + 1) = \log E_{min} + \log R_0' \tag{16}$$

23

where the logarithm has been taken to facilitate plotting on a
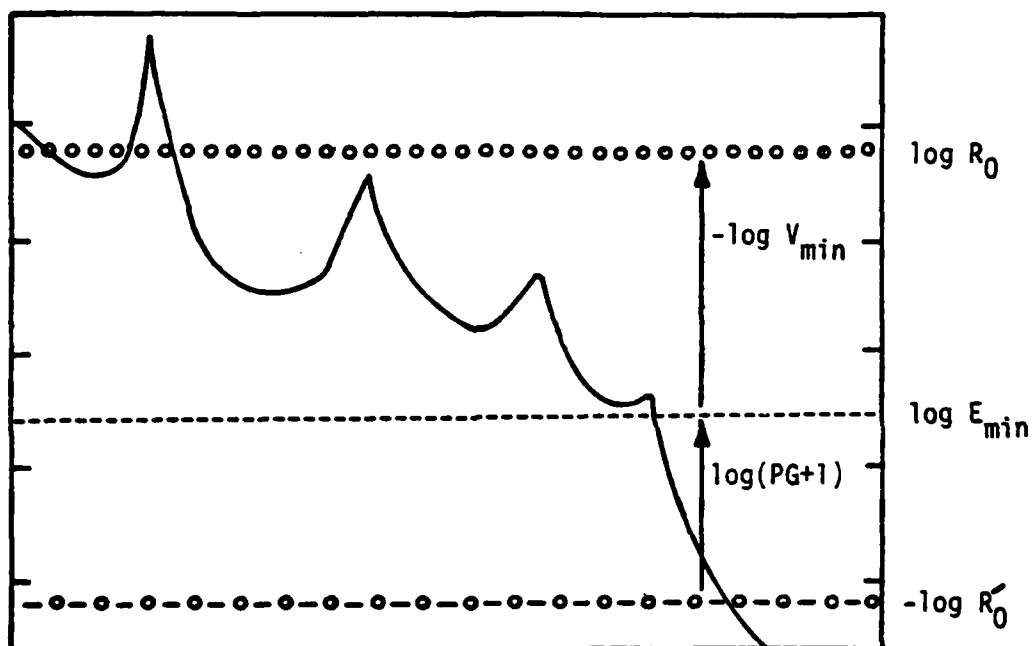decibel scale in Fig. 6.

FIG. 6. POWER AND PREDICTION GAINS OF LINEAR PREDICTION FILTER

From Fig. 6, it is clear that the power gain is related to
the inverse of the speech spectrum in the same way that the
prediction gain is related to the speech spectrum. In the next
section, a method for the reduction of the power gain with
minimal reduction of the prediction gain is investigated.

2.5 Schemes for Reduction of Power Gain

In previous sections, it was shown that a large power gain in the APC feedback loop can reduce the S/Q in the system. This was explicitly shown in (4) for a system with constant bit rate. When the power gain is nearly equal to the W/Q, this reduction in S/Q can be significant. It is often possible to reduce the PG by use of a suboptimal prediction filter in place of A(z) such that the system S/Q is increased. This happens when the resultant increase in the R/Q due to the lower PG is greater than the loss in S/R due to the suboptimal prediction filter.

Fig. 6 shows that it is the arithmetic mean of the inverse of the feedback filter spectrum that is important. This is primarily determined by the low energy portions of the speech spectrum, usually at the high frequencies. The method of high frequency correction attempts to modify the feedback filter by boosting the low energy parts of the speech spectrum before the linear prediction process. Since the power gain is very dependent on the low energy parts of the filter spectrum, i.e., the high energy sections of the inverse of the filter spectrum, and the prediction gain has only small dependence on those low energy sections, the power gain can be lowered with only a small reduction in the prediction gain.

In this quarter, the method of high frequency correction, introduced in the last report, was improved in an effort to maximize the S/Q. The basic method is to modify the autocorrelation vector representation of the speech signal used in the linear prediction analysis as shown by

$$R = R_{speech} + \lambda R_o V_p H \tag{17}$$

$$H = [0.375, -0.25, 0.0625, 0, 0, 0, \ldots 0] \tag{18}$$

where R is the new autocorrelation vector, R_{speech} is the original autocorrelation vector of the input speech signal $s(n)$, λ is a variable that determines the amount of modification, R_0 is the energy in the input speech frame, V_p is the normalized prediction error, and H is the autocorrelation vector of the impulse response of a high pass filter with two real zeros at $z=1$ in the z-plane. The scaling by $R_0 V_p$ which is equal to E_p, an approximation of the residual energy, assures that only low energy sections of the speech spectrum will be modified. The new filter found by linear prediction using the modified autocorrelation vector is suboptimal, having a lower prediction gain, and is a function of λ.

An increase in λ will cause both the power gain and the prediction gain to decrease. A reduction of the power gain will increase the R/W while a reduction of the prediction gain is a decrease in the S/R. The S/Q, proportional to the product of S/R and R/W, will have a maximum value for some value of positive or zero λ. The desired effect of the high frequency correction method is to optimize the S/Q. Experimental results show that the reduction in the power gain will cause the S/Q to increase in those frames that had large PG.

Since the relation of λ to the S/Q is different for every filter A(z), an iterative approach was used to optimize the system. The simplest scheme would be to choose a value of λ, compute the resultant filter, process the frame of speech, and then calculate the S/Q. A new value of λ could then be chosen and the process iterated until it converged to the optimum value. Since it is computationally expensive to calculate the APC loop for the frame of speech samples for each iteration, a method was developed to approximate the optimum value without that computation.

From Fig. 5, it is seen that small power gains result in R/W near 0 dB. The associated loss in S/Q is negligible. As the PG approaches W/Q, R/W decreases rapidly resulting in large decrease

of S/Q. Experiment results for a system using 2.12 bits per sample for coding, equivalent to a W/Q of approximately 11 dB, show that a decrease in the power gain to 3.6 (5.6 dB) will usually increase S/Q. Attempts to lower the power gain below 3.6 generally result in loss of S/Q due to the accompanying loss in prediction gain.

The iterative method first calculates the optimal filter by normal methods. If the filter power gain is larger than 3.6, a small value of λ is chosen, the autocorrelation vector is modified, the new filter's power gain is calculated. If the PG is still too large, the value of λ is increased until that power gain is close to 3.6.

For a typical utterance, the average S/Q of each frame increased 3.75 dB from the S/Q with no high frequency correction and 1.17 dB from the non-iterative method presented in the last report.

3. COMPUTATIONAL EFFICIENCIES FOR RESAMPLING

In the last report, it was noted that a major portion of the computation for the algorithm was due to the resampling operations. The input speech has been filtered previous to the analog to digital conversion process such that it does not contain energy at frequency components over 3.3 kHz. Thus, it is not necessary to sample the signal at the rate of 8 kHz. By reducing the sampling rate of the input speech from 8 kHz to 6.67 kHz for processing, the average number of bits for coding of the samples is increased from 1.80 to 2.16 bits per sample. After the speech is resynthesized, the signal is upsampled to 8 kHz. Unfortunately, the filtering involved in the resampling processes to avoid aliasing is computationally expensive.

The first method investigated for the reduction in the amount of computation is the use of different filters for the resampling process. Other methods will be reviewed in the future.

3.1 Finite Impulse Response Filters

The reference filter presently in use in the APC system is an equal-ripple finite impulse response (FIR) filter of length 250. All ripples in the spectrum of the pass band are of equal

amplitude as are the ripples in the filter stop band. The transition band extends from 3.12 kHz to 3.42 kHz. The minimum attenuation in the stop band is 50 dB. It has been shown previously that the resampling from 8 kHz to 6.67 kHz before coding and from 6.67 kHz to 8 kHz after resynthesis using this filter introduces no audible degradation into the system.

The trade-off for the lowered computational complexity of a shorter filter is the aliasing and/or attenuation of the high frequencies due to less attenuation in the stop band and/or a larger transition band. By using the properties of the speech signal and the properties of the analog filters used before and after the digitization processes, it is possible to use a shorter FIR filter without the causing audible degradations.

In general, the speech spectrum for voiced sounds decreases in magnitude as a function of increasing frequency. Distortions introduced at the high frequencies with energy proportional to the signal energy at those high frequencies may not be audible due to masking effects. Thus, some aliasing of those high frequencies may be permissible and, therefore, an FIR filter of length shorter than 250 may suffice.

If the speech signal is filtered by an analog filter before

digitization such that there is no signal energy above 3.3 kHz, there will be no aliasing due to the resampling down to 6.67 kHz if the digital filter used in the resampling passes no energy above 4.7 kHz. If the same analog filter is used after the digital to analog conversion, it will remove all aliasing due to resampling from 6.67 kHz to 8 kHz using the same digital filter.

For the digital resampling filter, a Hanning window FIR filter design algorithm was employed. This filter design was chosen because of the continued increase of attenuation in the stop band as a function of increasing frequency for low pass filters. The 64 point FIR filter has a maximum ripple of 0.05 dB in the passband and a minimum attenuation of 43.9 dB in the stopband. The pass band edge was 2.64 kHz and the stop band edge was 4.56 kHz. The filter frequency response was -3 dB at 3.33 kHz, -6 dB at 3.60 kHz, -10 dB at 3.83 kHz, and -20 dB at 4.10 kHz.

The performance of the filter in the resampling routine was evaluated by listening tests. Twelve utterances, sampled at 8 kHz, with SNR of 30 dB were downsampled to 6.67 kHz and then upsampled back to 8 kHz. No differences were perceived between those utterances processed with the 64 point filter and with the 250 point reference filter. Thus, the 64 point filter will

replace the 250 point filter for use in the APC system.   This
represents a factor of 4 savings in computation.

3.2   Infinite Impulse Response Filters

Infinite impulse response (IIR) filters are often much more
computationally efficient than FIR filters because of the added
flexibility of implementing poles and zeros.   In the special case
where the filter output will be downsampled, i.e., not every
output point will be used, the FIR filter may have an advantage.
In this section, the length of the IIR filter that requires the
same amount of computation as an N point FIR filter for use in
this resampling scheme will be investigated.

When not every point at the output of an FIR filter is used
because of a resampling operation to a lower rate, a
computational savings may be achieved by only computing the
output for those points that will be used.   IIR filters, however,
can not take advantage of this resampling each output point that
is needed is a function of several previous output points
including, in general, those output points that will be discarded
by downsampling.   For the APC system, this reduces the
computation of the FIR filters by a factor of 6 for resampling to
6.67 kHz and 5 for upsampling back to 8 kHz.

A linear phase FIR filter of order N, N even, requires N/2 multiplies per output point. The 64 point FIR filter resampled down by a factor of L will thus require 32/L multiplies per input point. An elliptic filter of order n will require (3n+3)/2 multiplies per input point [Rabiner & Gold, 1975], independent of output resampling. For L=5, the elliptic filter must be of order 3 or less to be computationally more efficient than the FIR filter of order 64. Tables show that elliptic filters of order 3 will not have as good specifications as the FIR filter [Rabiner & Gold, 1975]. Thus, for equivalent specifications, FIR filters will require less computation than elliptic filters.

4. CONCLUSIONS AND PLANS FOR FURTHER WORK

In Section 2, the problem of instability of the APC feedback loop due to excessive power gain of the feedback filter was studied. The iterative method of high frequency correction has been shown experimentally to increase the system signal to noise ratio and eliminate the "glitches" and "beeps", frames with extremely large encoding error.

Section 3 represented the beginning of work to reduce the computational complexity of the APC system. Use of the 64 point filter presented in that section reduces the computation by a factor of 4 over previous filters. In our future work, we will continue to examine methods to further reduce the computational complexity of the system. Topics to be studied include the processing of the 8 kHz data without downsampling and backward adaptation of the system parameters.

Since the quality of the coded speech may degrade with the use of schemes with reduced computational load, we will continue our efforts to improve speech quality. Such efforts include the implementation and testing of a pitch loop in the APC system and reducing the output noise power in the band 0 to 300 Hz since no speech is originally present in that band.

5. REFERENCES

[1] Rabiner, L.R. and Gold B., <u>Theory and Application of</u>
     <u>Digital Signal Processing</u>, Prentice-Hall, Inc., Englewood
     Cliffs, NJ., pp. 284-292, 1975.

DATE
FILME
8—8